FAKE NEWS DETECTION USING MACHINE LEARNING

NAME:HIBA NAFEESATH

ROLL.NO:20

PRODUCT OWNER:SYED FEROZE AHAMED M

Table Of Contents:

DESCRIPTION:

In traditional news making procedures, very limited and authorized individuals are involved and newspapers,radio,television were the only source of news. Due to these reasons news, credibility and authenticity are preserved. But in the era of internet, social network is becoming a news source of news. Easy and free access to these social networks makes the task of fabricating fake news and manipulating news a very effortless task. There is no authorize control point of these manipulated fake news which creates a question over there credibility and authenticity. The ease of getting direct news from the platform they mostly use has attracted the user. The reason to spread fake news can be social, political, and economical.

The rate at which it spread is very fast due to which controlling the spread manually is not possible.There is no platform via which the user can check the credibility and authenticity of the news and where authorities can directly inform about the fake news prevailing. Due to which people can believe in the news which can be a trouble for them and as well for society also. In the existing system, the action is taken after the adverse impact had already hit society. The proposed platform is useful for both common people and official authorities to prevent the spread of rumours in form of news. There is no existing platform that can verify the news and categorize it. This paper proposes a system that can be used for real-time prediction of news to be real or fake. This system is based on natural language processing to extract features from the data and then these features are used for the training of machine learning

This system is to detect whether the news is fake or not by using Cosine similarity Algorithm in machine learning.when we give news it will take words and similarities from that & compare it with the words in the internet,based on this the system will detect whether the news is fake or not.

MODULES:

1.Data collection

2.Data Cleaning

3.Data Preprocessing

4.Training

5.Testing

6.Result Generation

DEVELOPING ENVIRONMENT:

- Hardware specification:

  Processor : Intel Pentium Core i3 and above

  Primary Memory : 4 GB RAM and above

  Storage : 500 GB hard disk and above

- Software specification:

  Language :Python

  Front end : Python Django

  Back end : SQLite

  Operating system : windows 7 and above

  IDE : Visual Studio code,Jupyter Notebook

  Others : HTML,CSS

  Algorithm:Cosine Similarity Algorithm

  Technique:Natural Language Processing

  Data set:Fake news & True news from Kaggle website

PRODUCT BACKLOG:

User Story ID	Priority<High /Medium/Low >	Size(Hours)	Sprint	Status<Planned/Inprogress/ Completed>	Release Date	Release Goal
1	Medium	2	1	Completed	27/12/2021	Collection of datasets
2	High	3		Completed	28/12/2021	Preprocessing of collected data
3	Medium	3	2	Planned	29/12/2021	Visualisation of input data
4	High	2		Planned	15/01/2022-16/01/2022	Count no. of fake news & true news
5	Medium	5		Planned	22/01/2022	Split data into training & testing set
6	High	5		Planned	23/01/2022	Train the data
7	High	10	3	Planned	26/01/2022, 29/01/2022	UI designing(a form to enter news for checking)
8	High	20	4	Planned	05/02/2022-06/02/2022	Testing & collect input from user
9	High			Planned	12/02/2022	Generate Result

USER STORY:

User Story ID	As a <type of user>	I want to <perform some task>	So that I can <achieve some goal>
1	User	Collection of Dataset	Fake news Dataset
2	User	Preprocessing of collected data(null values elimination)	Cleaned final dataset
3	User	Visualisation of input data	Graphical representation of data
4	User	Count no. of fake news & true news	Number of datas
5	User	Split data into training & testing set	70% -training data & 30%-testing
6	User	Train the data using TFIDF vectorizer	Trained data
7	User	UI designing(a form to enter news for checking)	A form to enter news
8	User	Collect input from user & Testing	input from user & Tested data
9	User	generate output	Result(using Cosine similarity algorithm)

PROJECT PLAN:

User Story ID	Task Name	Start Date	End Date	Days	Status
1	Sprint 1	27/12/2021	27/12/2021	2	Completed
2		28/12/2021	28/12/2021		Completed
3	Sprint 2	29/12/2021	29/12/2021	5	Planned
4		15/01/2022	16/01/2022		Planned
5		22/01/2022	22/01/2022		Planned
6		23/01/2022	23/01/2022		Planned
7	Sprint 3	26/01/2022	29/01/2022	3	Planned
8	Sprint 4	05/02/2022	06/02/2022	3	Planned
9		12/02/2022	12/02/2022		Planed

SPRINT PLAN:

Backlog item	Status & Completion date	Original Estimate in hours	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13
User Story#1,2		Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours
Data collection	27/12/2021	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Preprocessing	28/12/2021	3	0	3	0	0	0	0	0	0	0	0	0	0	0
User story #3,4,5,6															
Visualisation & Training	23/01/2022	15	0	0	3	3	3	3	3	0	0	0	0	0	0
User story #7															
UI Designing	29/01/2022	10	0	0	0	0	0	0	0	4	4	2	0	0	0
User story #8,9															
Testing	12/02/2022	20	0	0	0	0	0	0	0	0	0	0	8	6	6
Total		50	2	3	3	3	3	3	3	4	4	2	8	6	6

Sprint1 Actual:

Backlog item	Status & Completion date	Original Estimate in hours	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10	Day 11	Day 12	Day 13	Completed<Y/N>
User Story#1,2		Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours	Hours
Data collection	27/12/2021	2	2	0	0	0	0	0	0	0	0	0	0	0	0	Y
Preprocessing	28/12/2021	3	0	3	0	0	0	0	0	0	0	0	0	0	0	Y
User story #3,4,5,6																
Visualisation & Training	23/01/2022	15	0	0	3	3	3	3	3	0	0	0	0	0	0	N
User story #7																
UI Designing	29/01/2022	10	0	0	0	0	0	0	0	4	4	2	0	0	0	N
User story #8,9																
Testing	12/02/2022	20	0	0	0	0	0	0	0	0	0	0	8	6	6	N
Total		50	2	3	3	3	3	3	3	4	4	2	8	6	6	

Thank you